Exploring the Relationship Between Data Science and Circular Economy: An Enhanced CRISP-DM Process Model

Eivind Kristoffersen[1(✉)], Oluseun Omotola Aremu[2], Fenna Blomsma[3], Patrick Mikalef[1], and Jingyue Li[1]

[1] Norwegian University of Science and Technology, 7491 Trondheim, Norway
eivind.kristoffersen@ntnu.no
[2] The University of Queensland, Brisbane, QLD 4072, Australia
[3] Technical University of Denmark, 2800 Kgs Lyngby, Denmark

Abstract. To date, data science and analytics have received much attention from organizations seeking to explore how to use their massive volumes of data to create value and accelerate the adoption of Circular Economy (CE) concepts. The correct utilization of analytics with circular strategies may enable a step change that goes beyond incremental efficiency gains towards a more sustainable and circular economy. However, the adoption of such smart circular strategies by the industry is lagging, and few studies have detailed how to operationalize this potential at scale. Motivated by this, this study seeks to address how organizations can better structure their data understanding and preparation to align with overall business and CE goals. Therefore, based on the literature and a case study the relationship between data science and the CE is explored, and a generic process model is proposed. The proposed process model extends the Cross Industry Standard Process for Data Mining (CRISP-DM) with an additional phase of *data validation* and integrates the concept of *analytic profiles*. We demonstrate its application for the case study of a manufacturing company seeking to implement the smart circular strategy - predictive maintenance.

Keywords: Data science · Circular Economy · Predictive maintenance · Business analytics · CRISP-DM

1 Introduction

In recent years, the concept of Circular Economy (CE) has received significant attention from businesses, policymakers, and researchers as a way to promote sustainable development [25]. With the aim of decoupling value-creation from the consumption of finite resources, CE leverages a range of restorative, efficiency, and productivity oriented strategies to keep products, components, and materials in use for longer [16,17]. Nevertheless, the adoption of CE by the industry so far is modest [26,54,61]. This also holds for manufacturing companies. Although

Published by Springer Nature Switzerland AG 2019
I. O. Pappas et al. (Eds.): I3E 2019, LNCS 11701, pp. 177–189, 2019.
https://doi.org/10.1007/978-3-030-29374-1_15

they play a vital role in the creation of value, little improvements are seen in their decoupling from linear consumption of resources.

In parallel, the emergence of new technologies as the Internet of Things, Big Data, and Artificial Intelligence - collectively known as Digital Technologies (DTs) - have encouraged a paradigm shift for industrial production, the 'Fourth Industrial Revolution'. These DTs are seen as one of the key enablers for a wider adoption and accelerated transition to CE [19,20]. Moreover, they form the operational building blocks of a more efficient and effective CE, the Smart CE.

The significance of DTs to transition to a CE however is argued to be more than a technical challenge [64]. First, it requires a clear data and business analytics strategy, the right people to effect a data-driven cultural change, and it demands the organization to appropriately structure their departments to align the analytics capability with their overall business strategy. Kiron and Shockley [36], concur and note that organizations have to develop data-oriented management systems both to make sense of the increasing volumes of data and, more importantly, for transforming the insights into business value and a competitive advantage. Supporting this transformation, by the use of analytics methods, is the data science process[1] [57]. However, there seems to be a gap between the output of these insights and the generation of business value [14,44,66]. As highlighted by extensive research, this is often due to the ineffective integration of data science methods within the organization [2,14,21,38,66].

Extant data science methodologies have not yet been scoped or demonstrated for the context of CE. For instance, the study [20] only presents the need for a process covering data collection, data engineering, algorithm development, and algorithm refinement within the CE without detailing how to operationalize it. Contributions are more commonly seen on topics such as service design [45], or the technical details of analyzing data, e.g., [11]. In this work, we recognize the importance of aligning an organizations analytics development with overall business and CE initiatives. The process discussed in this paper differs from previous contributions in three ways: First, it extends the Cross-Industry Standard Process for Data Mining (CRISP-DM) with an additional phase of *data validation*. Second, it consolidates an organization's analytics knowledge base by integrating the concept of *analytic profiles*. Third, the process is demonstrated for the context of CE by the case study of predictive maintenance (PdM) for an original equipment manufacturer (OEM). We use PdM as an example here as it is a prominent smart circular strategy (facilitating for extending the use-cycle, increasing the utilization and looping/cascading assets), allowing for generalization to other strategies.

The remainder of the work is detailed in following sections. Section 2 gives background on the data science and the concept of CE, thereafter Sect. 3 presents the research approach followed for this work. Section 4 presents the proposed CRISP-DM process model modifications, whilst Sect. 4.1 details the case study of PdM for CE. Finally, the paper is concluded and further work presented in Sect. 5.

[1] In this paper, we use the expressions process, method, and methodology interchangeable as a set of activities that interact to produce a result.

2 Background

2.1 Data Science

Data science is a multidisciplinary field encompassing tools, methods, and systems from statistics and data analytics (hereby referred to as analytics) applied to large volumes of data with the purpose of deriving insights for decision-making support [21,38,48,57,66]. As such, data science may include the collection and use of data to: *(i)* better understand the business operation and provide current state evaluation of performance, *(ii)* transform the organization from being reactive to proactive in business decision-making through use of predictive analytics, *(iii)* improve customer service through use of data to build a more coherent knowledge base and understanding of customer needs, and *(iv)* increase the efficiency, enhance the effectiveness and facilitate the implementation of CE concepts at scale (e.g., by optimizing circular infrastructures, business models, and products-service systems) [13,20,44,47,48].

Fig. 1. Phases of the CRISP-DM process model [10]

Research shows that companies embracing data science have experienced noticeable gains in business development (i.e., productivity and profitability) [44,66]. However, the impact of data science is not limited to commercial endeavours alone. For instance, studies show improved sustainability for building energy management [46], predictive capabilities in supply chain management [66], health care services in the medical industry [50] and environmental impact of the manufacturing and process industry [29,34]. However, the effects for the CE is still largely unexplored.

To support the effective integration of data science within organizations, various methodologies have been proposed in the literature (e.g., KDD and SEMMA [22,59]). The most commonly used is the CRISP-DM process model created by IBM, reporting a use level of 43% followed by 28% of companies using their own methodology [53]. CRISP-DM is described in terms of a hierarchical and cyclic process model composed of six phases (see Fig. 1), each consisting of several generic tasks (e.g., clean data), specialized tasks (e.g., cleaning of numerical and categorical values) and process instances (i.e., how these tasks are operationalized through different actions, decisions and results). The methodology is designed to be generic, complete and stable, meaning that it should cover the whole analytics development process for all possible applications, and should be valid for yet unforeseen developments (e.g., new analytics modeling techniques) [10]. Despite the high reported level of use, the methodology appears to not be in active development. We recognize that IBM have later proposed an extension to CRISP-DM called the Analytics Solutions Unified Method (ASUM-DM) [30]. However, ASUM-DM differs only in the operational/deployment aspects of the process and describes the same phases for development. Therefore, given CRISP-DM's continued widespread adoption from practitioners and inherent generic, complete and stable design, we have chosen it as our reference model. As a stand-alone data science process, CRISP-DM has been successful within its bounds [67]. However, suggestions for the following shortcomings have been made [6,55] (the issues are addressed in Sect. 4):

(i) the lack of a good management view to track and communicate knowledge/insights,
(ii) the lack of assessment of analytics implementation feasibility (e.g. by leveraging a maturity assessment or gap analysis),
(iii) despite its widespread adoption, the process is not always understood by the wider business community, hence it is difficult to manage actual business value of the analyses,
(iv) the iterations do not loop back to the business level (prior to analytics modeling) for domain specific knowledge after the first two phases,
(v) and lack of control of added value.

2.2 Circular Economy

CE emerged as an umbrella concept in the 2010s as an approach to achieve sustainability [7], and encompass a range of strategies for narrowing, slowing and closing material and energy flows [8,18] as a means for addressing structural waste. Although the CE concept continues to grow and gain attention, it remains in an early stage of development. Therefore, a detailed definition of CE is still missing in the literature [24,31,35,41]. However, one of the most prominent definitions has been provided by the Ellen MacArthur Foundation [15,17], where CE is defined as a system *"that provides multiple value creation mechanisms, which are decoupled from the consumption of finite resources."*

CE strategies span from operational processes (i.e., restore, reduce, recirculate, and avoid) to more strategic, and business models related, strategies (i.e., reinvent, rethink, and reconfigure). DTs is highlighted by literature as an

important enabler of CE strategies [4,9,19,49,51]. However, the adoption by industry is meager, and the research is still in a pre-paradigmatic stage [51]. Using DTs for the CE, Smart CE, promotes a sustainable ecosystem where assets (products, components, materials, and so on) are given virtual, or digital counterparts that allows for the sensing, communication, interaction, and exchange of data. By embedding software and analytics intelligence within or connected to these assets allows for easier manipulation and automation of the assets and of the environment, or system, in which they operate - enabling an increase of the systemic resource efficiency and productivity of the CE. This can for instance be seen with the data-driven maintenance strategy, or smart circular strategy, PdM [1,43,62]. PdM is a pertinent strategy for OEMs seeking to transition to the CE. OEMs offer one of the highest potential for environmental and economic impact of any sector [19]. In the European Union, material savings alone have been estimated to USD 650 billion for a full CE transition [15]. A gross part of this potential can be linked back to PdM by its three CE value drivers [19]:

Extending the life cycle: correct condition-assessment for need of and scheduling of appropriate life cycle extending operations,
Increasing utilization: reduce unplanned downtime and increased equipment effectiveness,
Looping the asset: improve insight and transparency into asset's condition and usage history.

Achieving a Smart CE requires companies to reconfigure and blend their existing value creation mechanisms with new innovative digital strategies. Blending digital strategies with value offerings require companies to become data-driven (i.e., decision-makers base their actions on data and insights generated from analytics, rather than instinct). Supporting this, Janssen et al. [33] argue that the quality of these evidence-based decisions depends largely on the quality of the inputs and the process that transforms these inputs into outputs - essentially the data science process.

3 Research Approach

The proposed process was developed based on an analysis of the *data understanding* and *data preparation* phases of the current CRISP-DM 1.0 step-by-step data mining guide [10] together with insights from company engagement under the CIRCit research project [12]. Given the exploratory nature of the research and the pre-paradigmatic stage of the field [51], case study research was chosen as the methodology for empirical investigation [69]. The case study research methodology is particularly suitable for the initial stage of investigation [31] as it help provide insights with relatively good understanding of the complexity and nature of the phenomenon [65]. Moreover, even a single case study can provide scientific development through a deep understanding of the problem and the capturing of experiences [23].

A research protocol was used in order to ensure reliability and validity of the findings, including case study design, data collection, data analysis, and formalization of results [69]. The company was selected based on a judgmental sampling

technique [28]. First, the company should be from the manufacturing industry and have interest in, or experience with, the CE. Second, the company need to have sensory/operation data available for analytics and Smart CE investigation for this paper. To this regard, a Nordic OEM company manufacturing and servicing industrial cranes, who is particular interested in PdM, was contacted and accepted to participate in the project and case study. However, the company identity has been concealed here to protect their business interests.

Following the research protocol, data collection was performed through several semi-structured interviews to first gather general information about the context of the company before the operation data were exchanged and insights specific to analytics and PdM were collected. Following the collection of organizational and operation data, analytics investigation was performed to evaluate the potential PdM and set implementation requirements. Then, the last face of the protocol was conducted, looking for possible procedural improvements of the CRISP-DM model to meet the requirements from analytics.

4 An Enhanced CRISP-DM Process Model

Asset and process management research argue that data should be specifically structured for the intended use within the work flow [27,57]. Analytics research concur and note that insight is more obtainable when the data has been preprocessed for a specific domain of analysis [32,37,42,52,68]. To this effect, and to address the previous highlighted shortcomings, we propose an extended CRISP-DM process model. The proposed process model adds an additional phase called *data validation* (addressing issues (iv) and (v)), and argues for the integration of *analytic profiles* (addressing issues (i) and (iii)) as a core element of the process. Figure 2 illustrates the enhanced CRISP-DM process model developed.

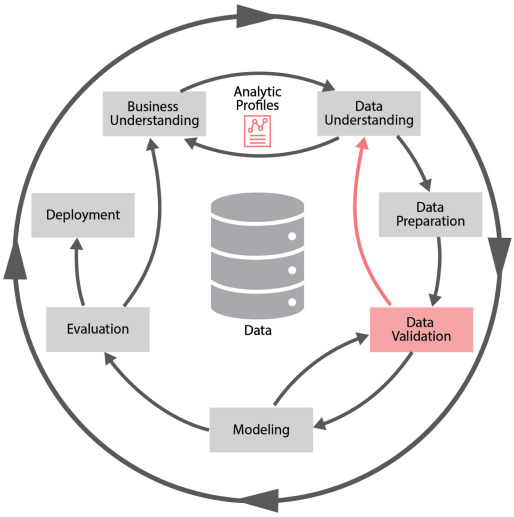

Fig. 2. An enhanced CRISP-DM process model

In CRISP-DM, there is no validation between the *data preparation* phase and the *modeling* phase against the specific business domain [6,48]. Specifically, once the data is prepared for modeling, only the criterion needed to ensure optimal analytics model performance are considered [48,67]. Thus, a complete understanding of whether the data which is prepared is a valid representation of the original problem is not guaranteed. General data preparation methods alter the original data, and there is often loss in information specific to the domain that should be monitored [5,48]. As such, this may result in sub-optimal solutions that miss the mark on the intended capturing of business value [55,63]. Therefore, we argue that data validation should be done by the re-involvement of the business entity, or domain experts, to validate that a proper understanding of the data and business problem have been reached, and include data preparation methods tailored for the given analytic profile. The data validation phase may result in a re-iteration of the data understanding and/or the data preparation phase(s) (indicated by a single arrow back in the diagram).

Analytic profiles are defined as structures that standardize the collection, application and re-use of analytics insights and models for key business entities [60]. As such, an analytic profile is an abstract collection of knowledge, mainly used in the business and data understanding phases, that lists the best practices for a particular analytics use case, or problem. Analytic profiles may have different levels of granularity depending in the use case and the organization's level of experience. However, information on the following elements should be included:

- **Use case description** defining the business goal (e.g., predict the remaining useful life of a crane),
- **Domain specific insights** important for the use case (e.g., knowledge about typical crane failures and causes),
- **Data sources** relevant for the use case (e.g., time-series data of crane operation and service data with failure modes),
- **Key Performance Indicators (KPIs)** or metrics for assessing the analytics implementation performance (e.g., crane failure rate, downtime and maintenance costs),
- **Analytics models and tools** with proven conformity for the given problem (e.g., long short-term memory networks and deep belief networks),
- **Short descriptions of previous implementations** with lessons learned (e.g., deep belief networks for backlash error prediction in machining centers [40]).

As per the CRISP-DM process level breakdown [10], analytic profiles can be regarded as a generic task particularly relevant between the business and data understanding phases (indicated by an analytic profile icon in the diagram). Through such a consolidation of the analytics knowledge base, organizations can more easily learn and reuse their own experience and the experience of others to catalyze the analytics development process. Furthermore, Kiron and Shockley [36] state that organizations should appropriately structure their resources to

align their analytics capability with their overall business strategies. Therefore, we argue that analytic profiles should be build for all business strategies, or use cases, relying on insights from analytics.

4.1 Case Study: Predictive Maintenance for an Original Equipment Manufacturer

In this section we give detail to the strategy of PdM for the context of CE together with insights from the case study to validate the adaptations made to CRISP-DM. However, we only detail the structuring of data from the data understanding phase to the data validation phase. As such, we do not cover the whole analytics development process or the full contents of the analytic profile of PdM.

According to EN 13306:2010, predictive maintenance is defined as condition-based maintenance carried out following a forecast from analytics or known characteristics of the features of the degradation of an asset. It contrasts traditional, or non-predictive, maintenance actions that are only based on information of the current condition. Therefore, as PdM integrates multiple DTs (e.g. Internet of Things and Artificial Intelligence) it enables real-time access to detailed information about the assets' location, condition, and availability. This allows for augmenting human decision-making by predicting product health, wear, usage, and energy consumption [56]. This *"sense and respond"* capability is crucial for the CE as it allow for greater transparency of assets' actual condition throughout their life cycle, and enable triggering of appropriate life cycle extending operations for the OEM or service provider [58].

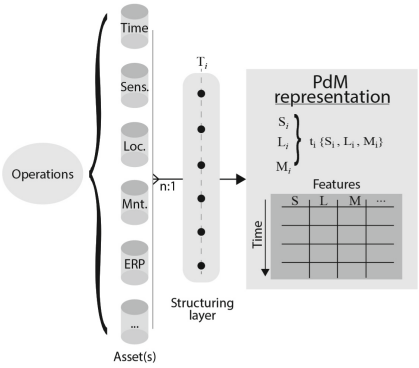

Fig. 3. Example structuring of data for a PdM analytic profile

The main goal of the analytics exploration was to evaluate the current status of analytics development towards the implementation of PdM within the company. For the case of a PdM analytic profile, the occurrence of faults or

degradation and their influence during assets' life cycle are considered domain specific knowledge [3,5]. Therefore, the data must contain life cycle observations in which information or knowledge pertaining to the occurrence of faults, degradation, or process change can be inferred [39,52,62]. In general, this can be decomposed to sensor measurements S, location L, and maintenance logs M which describe the condition at various time steps. Figure 3 illustrates such a structuring of an asset's data in which its attributes are collected from multiple data sources, such as time, sensory/monitoring data, location, maintenance logs, and Enterprise Resource Planning (ERP) system data. The observation at an arbitrary time t_i describes the condition of the asset per set of attributes $t_i(S_i, L_i, M_i)$. This structuring ensures the data is useful for the intended analysis, and when combined with involvement from the business entity by domain experts makes up the *data validation* phase. However, the analytics exploration performed by the researchers showed that the current collected features were not sensitive enough to the failure categories required by PdM. This means that the provided data lacked in quality and did not contain the necessary level of detail of failure modes needed in order to predict impending failures. Consequently, the business goal and targeted analyses had to be changed to less advanced analyses. In this case, the goal was transferred to abnormality identification and the development of a method to evaluate the severity degree of the cranes. High severity degree means that the behaviour of the sample crane is different from the majority, thus is more likely to have impending failures. Also, it is not uncommon that important information, or observations, within the data might get 'lost', or disregarded, in the data preparation phase (due to misunderstanding of the business goal). Therefore, we argue that it is crucial for the success of data science initiatives to include a phase of data validation prior to modeling. In summary, the data validation phase ensures that modeling happens on the right data for the right reasons.

Following the data preparation and data validation phases, the standard CRISP-DM phases of *modeling*, *evaluation*, and *deployment* should be followed. In these phases, analytics methods are applied to, e.g., provide predictions or current state inferences of the manufacturing operation. This may include the accurate identification and prediction of impending failures, degradations, or abnormal behaviour, which can then be used for decision-making support or directive actions for operations management. Finally, when the process of PdM has been structured in such a way that it allows for standardized collection, application and re-use of its analytics insights.

Interviews with the case company revealed that such a structuring of the data and standardized use of analytic profiles had not been systematically integrated within the organization. In the intervention after the analytics exploration the researchers presented the results of their analyses with suggestions for how to appropriately structure their data science process model (e.g., how to link the abnormality identification with typical uses cases and KPIs). Feedback from the company showed the new data science process, especially with the active use of KPIs, could provide a better management view for easier communication of knowledge, tracking of business value and CE impact.

5 Conclusion and Future Work

This paper proposed an enhanced CRISP-DM process model and a case study discussing how to structure the data of the analytic profile of PdM for the context of CE. We addressed the issues (iv) and (v) (lack of iterations looping back to the business level and no control of added value) by introducing an additional phase of data validation. As such, we highlighted the importance of the re-involvement of the business entity, or domain experts, to include domain specific knowledge for structuring and validating the data prior to modeling. Furthermore, we partly addressed the issues (i) and (iii) (lack of good management view and difficulty in managing actual business value of analyses) by introducing analytic profiles as an integrative part of the process model. Motivated by the benefits of the Smart CE, we discussed how data science is fundamental for using DTs to increase the efficiency, enhance the effectiveness and facilitate the implementation of CE strategies. For future work, we aim to extend on the business analytics and CE connection to the data science process. Essentially, detailing the business understanding and data understanding phases with CE related business model scoping and analytics leverage assessment. Lastly, greater detail and empirical evaluation of the suggested CRISP-DM modification should be added.

Acknowledgment. The authors would like to thank the reviewers of this paper. Also, the authors would like to acknowledge that this work was conducted as part of the research project CIRCit (Circular Economy Integration in the Nordic Industry for Enhanced Sustainability and Competitiveness), which is part of the Nordic Green Growth Research and Innovation Programme (grant number: 83144), and funded by NordForsk, Nordic Energy Research, and Nordic Innovation.

References

1. Alaswad, S., Xiang, Y.: A review on condition-based maintenance optimization models for stochastically deteriorating system. Reliab. Eng. Syst. Saf. **157**, 54–63 (2017)
2. Amankwah-Amoah, J., Adomako, S.: Big data analytics and business failures in data-rich environments: an organizing framework. Comput. Ind. **105**, 204–212 (2019)
3. An, D., Choi, J.H., Kim, N.H.: Prognostics 101: a tutorial for particle filter-based prognostics algorithm using Matlab. Reliab. Eng. Syst. Saf. **115**, 161–169 (2013)
4. Antikainen, M., Uusitalo, T., Kivikytö-Reponen, P.: Digitalisation as an enabler of circular economy. Proc. CIRP **73**, 45–49 (2018)
5. Aremu, O.O., Salvador Palau, A., Hyland-Wood, D., Parlikad, A.K., McAree, P.R.: Structuring data for intelligent predictive maintenance in asset management. In: 16th IFAC Symposium on Information Control Problems in Manufacturing (2018)
6. Bahrepour, M.: The forgotten step in CRISP-DM and ASUM-DM methodologies. https://sharing.luminis.eu/blog/the-forgotten-step-in-crisp-dm-and-asum-dm-methodologies/. Accessed 08 Mar 2019
7. Blomsma, F., Brennan, G.: The emergence of circular economy: a new framing around prolonging resource productivity. J. Ind. Ecol. **21**(3), 603–614 (2017)

8. Bocken, N.M., Short, S.: Towards a sufficiency-driven business model: experiences and opportunities. Environ. Innov. Soc. Transitions **18**, 41–61 (2016)
9. Bressanelli, G., Adrodegari, F., Perona, M., Saccani, N.: The role of digital technologies to overcome circular economy challenges in PSS business models: an exploratory case study. Proc. CIRP **73**, 216–221 (2018)
10. Chapman, P., Clinton, J., Kerber, R., Khabaza, H., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step-by-step data mining guide. Technical report (2000)
11. Cielen, D., Meysman, A., Ali, M.: Introducing data science: big data, machine learning, and more, using Python tools. Manning Publications Co. (2016)
12. CIRCit: Circit project page (2019). http://circitnord.com/
13. Dhar, V.: Data science and prediction. NYU Working Paper No. 2451/31635 (2012)
14. Domino Data Lab: Key factors on the journey to become model-driven - a survey report (2018)
15. MacArthur, E.: Towards a Circular Economy and Business Rationale for an Accelerated Transition. Ellen MacArthur Foundation, Cowes, UK (2013)
16. Ellen MacArthur Foundation: Delivering the circular economy: a toolkit for policymakers (2015)
17. MacArthur, E.: Growth within: a circular economy vision for a competitive Europe. Ellen MacArthur Foundation, Cowes, UK (2015)
18. Ellen MacArthur Foundation: Towards a circular economy: business rationale for an accelerated transition (2015). Accessed 25 Oct 2016
19. MacArthur, E.: Intelligent assets. unlocking the circular economy potential. Ellen MacArthur Foundation, Cowes, UK (2016)
20. MacArthur E.: Artificial intelligence and the circular economy. Ellen MacArthur Foundation, Cowes, UK (2019)
21. Elshawi, R., Sakr, S., Talia, D., Trunfio, P.: Big data systems meet machine learning challenges: towards big data science as a service. Big data research (2018)
22. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. Al Mag. **17**(3), 37 (1996)
23. Flyvbjerg, B., Budzier, A.: Why your it project may be riskier than you think (2011)
24. Geng, Y., Doberstein, B.: Developing the circular economy in China: challenges and opportunities for achieving 'leapfrog development'. Int. J. Sustain. Develop. World Ecol. **15**(3), 231–239 (2008)
25. Ghisellini, P., Cialani, C., Ulgiati, S.: A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. J. Clean. Prod. **114**, 11–32 (2016)
26. Haas, W., Krausmann, F., Wiedenhofer, D., Heinz, M.: How circular is the global economy: an assessment of material flows, waste production, and recycling in the European Union and the world in 2005. J. Ind. Ecol. **19**(5), 765–777 (2015)
27. Haddar, N., Tmar, M., Gargouri, F.: A framework for data-driven workflow management: modeling, verification and execution. In: Decker, H., Lhotská, L., Link, S., Basl, J., Tjoa, A.M. (eds.) DEXA 2013. LNCS, vol. 8055, pp. 239–253. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40285-2_21
28. Henry, G.T.: Practical Sampling, vol. 21. Sage (1990)
29. Ho, T.C., Mat, S.C.K.M.Z., San, L.H., et al.: A prediction model for CO2 emission from manufacturing industry and construction in Malaysia. In: 2015 International Conference on Space Science and Communication (IconSpace). pp. 469–472. IEEE (2015)
30. IBM: Analytics solutions unified method - implementations with agile principles (2016). ftp://ftp.software.ibm.com/software/data/sw-library/services/ASUM.pdf

31. Jabbour, C.J.C., de Sousa Jabbour, A.B.L., Sarkis, J., Godinho Filho, M.: Unlocking the circular economy through new business models based on large-scale data: an integrative framework and research agenda. Technol. Forecast. Soc. Change **144** 546–552 (2017)
32. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning, vol. 112. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-7138-7
33. Janssen, M., van der Voort, H., Wahyudi, A.: Factors influencing big data decision-making quality. J. Bus. Res. **70**, 338–345 (2017)
34. Kameswari, U.S., Babu, I.R.: Sensor data analysis and anomaly detection using predictive analytics for process industries. In: 2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI), pp. 1–8. IEEE (2015)
35. Kirchherr, J., Reike, D., Hekkert, M.: Conceptualizing the circular economy: an analysis of 114 definitions. Resour. Conserv. Recycl. **127**, 221–232 (2017)
36. Kiron, D., Shockley, R.: Creating business value with analytics. MIT Sloan Manage. Rev. **53**(1), 57 (2011)
37. Kun, W., Tong, L., Xiaodan, X.: Application of big data technology in scientific research data management of military enterprises. Proc. Comput. Sci. **147**, 556–561 (2019)
38. Larson, D., Chang, V.: A review and future direction of agile, business intelligence, analytics and data science. Int. J. Inf. Manag. **36**(5), 700–710 (2016)
39. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., Lin, J.: Machinery health prognostics: a systematic review from data acquisition to RUL prediction. Mech. Syst. Signal Process. **104**, 799–834 (2018)
40. Li, Z., Wang, Y., Wang, K.: A data-driven method based on deep belief networks for backlash error prediction in machining centers. J. Intell. Manuf. 1–13 (2017). https://doi.org/10.1007/s10845-017-1380-9
41. Lieder, M., Rashid, A.: Towards circular economy implementation: a comprehensive review in context of manufacturing industry. J. Clean. Prod. **115**, 36–51 (2016)
42. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing sax: a novel symbolic representation of time series. Data Min. Knowl. Disc. **15**(2), 107–144 (2007)
43. Liu, B., Liang, Z., Parlikad, A.K., Xie, M., Kuo, W.: Condition-based maintenance for systems with aging and cumulative damage based on proportional hazards model. Reliab. Eng. Syst. Saf. **168**, 200–209 (2017)
44. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D., Barton, D.: Big data: the management revolution. Harvard Bus. Rev. **90**(10), 60–68 (2012)
45. Meierhofer, J., Meier, K.: From data science to value creation. In: Za, S., Drăgoicea, M., Cavallari, M. (eds.) IESS 2017. LNBIP, vol. 279, pp. 173–181. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56925-3_14
46. Molina-Solana, M., Ros, M., Ruiz, M.D., Gómez-Romero, J., Martín-Bautista, M.J.: Data science for building energy management: a review. Renew. Sustain. Energy Rev. **70**, 598–609 (2017)
47. Nath, P., Nachiappan, S., Ramanathan, R.: The impact of marketing capability, operations capability and diversification strategy on performance: a resource-based view. Ind. Mark. Manag. **39**(2), 317–329 (2010)
48. Newman, R., Chang, V., Walters, R.J., Wills, G.B.: Model and experimental development for business data science. Int. J. Inf. Manag. **36**(4), 607–617 (2016)
49. Nobre, G.C., Tavares, E.: Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study. Scientometrics **111**(1), 463–492 (2017)

50. Ottenbacher, K.J., Graham, J.E., Fisher, S.R.: Data science in physical medicine and rehabilitation: opportunities and challenges. Phys. Med. Rehabil. Clin. N. Am. **30**(2), 459–471 (2019)
51. Pagoropoulos, A., Pigosso, D.C., McAloone, T.C.: The emergent role of digital technologies in the circular economy: a review. Proc. CIRP **64**, 19–24 (2017)
52. Peng, Y., Dong, M., Zuo, M.J.: Current status of machine prognostics in condition-based maintenance: a review. Int. J. Adv. Manuf. Technol. **50**(1–4), 297–313 (2010)
53. Piatetsky, G.: CRISP-DM, still the top methodology for analytics, data mining, or data science projects. In: KDD News (2014)
54. Planing, P.: Business model in novation in a circular economy reasons for non-acceptance of circular business models. Open J. Bus. Model Innov. **1**, 11 (2015)
55. Ponsard, C., Touzani, M., Majchrowski, A.: Combining process guidance and industrial feedback for successfully deploying big data projects. Open J. Big Data (OJBD) **3**(1), 26–41 (2017)
56. Porter, M.E., Heppelmann, J.E.: How smart, connected products are transforming competition. Harvard Bus. Rev. **92**(11), 64–88 (2014)
57. Provost, F., Fawcett, T.: Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media Inc, Newton (2013)
58. Romero, D., Noran, O.: Towards green sensing virtual enterprises: interconnected sensing enterprises, intelligent assets and smart products in the cyber-physical circular economy. IFAC-PapersOnLine **50**(1), 11719–11724 (2017)
59. SAS: Semma. https://www.sas.com/en_gb/software/analytics-overview.html. Accessed 10 Apr 2019
60. Schmarzo, B.: Big Data MBA: Driving Business Strategies with Data Science. Wiley, Hoboken (2015)
61. Sousa-Zomer, T.T., Magalhães, L., Zancul, E., Cauchick-Miguel, P.A.: Exploring the challenges for circular business implementation in manufacturing companies: an empirical investigation of a pay-per-use service provider. Resour. Conserv. Recycl. **135**, 3–13 (2018)
62. Susto, G.A., Schirru, A., Pampuri, S., McLoone, S., Beghi, A.: Machine learning for predictive maintenance: a multiple classifier approach. IEEE Trans. Ind. Inf. **11**(3), 812–820 (2015)
63. Viaene, S.: Data scientists aren't domain experts. IT Prof. **15**(6), 12–17 (2013)
64. Vidgen, R., Shaw, S., Grant, D.B.: Management challenges in creating value from business analytics. Eur. J. Oper. Res. **261**(2), 626–639 (2017)
65. Voss, C.: Case research in operations management. In: Researching operations management, pp. 176–209. Routledge (2010)
66. Waller, M.A., Fawcett, S.E.: Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. J. Bus. Logistics **34**(2), 77–84 (2013)
67. Wirth, R., Hipp, J.: CRISP-DM: towards a standard process model for data mining. In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, pp. 29–39. Citeseer (2000)
68. Wood, D., Zaidman, M., Ruth, L., Hausenblas, M.: Linked Data. Manning Publications Co. (2014)
69. Yin, R.K.: Applied social research methods series case study research: design and methods (1984)